

# Spatial Entropy Pursuit for Fast and Accurate Perimetry Testing

Derk Wild, Şerife Seda Kucur, and Raphael Sznitman

Ophthalmic Technology Laboratory, ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

Correspondence: Raphael Sznitman, Ophthalmic Technology Laboratory, ARTORG Center for Biomedical Engineering Research, University of Bern, Murtenstrasse 50, CH-3008, Bern, Switzerland; raphael.sznitman@artorg.unibe.ch.

Submitted: November 19, 2016

Accepted: May 22, 2017

Citation: Wild D, Kucur ŞS, Sznitman R. Spatial entropy pursuit for fast and accurate perimetry testing. *Invest Ophthalmol Vis Sci.* 2017;58:3414–3424. DOI:10.1167/iovs.16-21144

**PURPOSE.** To propose a static automated perimetry strategy that increases the speed of visual field (VF) evaluation while retaining threshold estimate accuracy.

**METHODS.** We propose a novel algorithm, spatial entropy pursuit (SEP), which evaluates individual locations by using zippy estimation by sequential testing (ZEST) but additionally uses neighboring locations to estimate the sensitivity of related locations. We model the VF with a conditional random field (CRF) where each node represents a location estimate that depends on itself as well as its neighbors. Tested locations are randomly selected from a pool of locations and new locations are added such that they maximally reduce the uncertainty over the entire VF. When no location can further reduce the uncertainty significantly, remaining locations are estimated from the CRF directly.

**RESULTS.** SEP was evaluated and compared to tendency-oriented strategy, ZEST, and the Dynamic Test Strategy by using computer simulations on a test set of 245 healthy and 172 glaucomatous VFs. For glaucomatous VFs, root-mean-square error (RMSE) of SEP was comparable to that of existing strategies (3.4 dB), whereas the number of stimulus presentations of SEP was up to 23% lower than that of other methods. For healthy VFs, SEP had an RMSE comparable to evaluated methods (3.1 dB) but required 55% fewer stimulus presentations.

**CONCLUSIONS.** When compared to existing methods, SEP showed improved performances, especially with respect to test speed. Thus, it represents an interesting alternative to existing strategies.

**Keywords:** static perimetry, perimetry testing, machine learning

Given its long history, visual field estimation by perimetry examination remains a standard approach for detection and monitoring of visual impairment. Static white-on-white automated perimetry (SAP) is perhaps the most common form and has been paramount for large clinical studies measuring visual function for glaucomatous patients<sup>1,2</sup> and patients with neurological disorders.<sup>3</sup>

At the heart of SAP lies a sequential light-stimulus and patient response exchange to estimate local “sensitivity thresholds” at specific locations of the visual field, that is, the amount of light stimulus that would induce a 50% chance of being identified by a subject. Typically used to measure the central 24° to 30° of the visual field, presenting all possible stimulus intensities at all locations, multiple times, would allow a highly accurate estimation of the complete visual field. Unfortunately, such a strategy would be desperately slow and exhausting for the subject. Conversely, presenting a single stimulus to estimate a handful of locations would be fast but highly inaccurate. As such, perimetry strategies implicitly suffer from an accuracy-speed trade-off, where the ideal goal is to be both fast and accurate.

Given this fact, numerous strategies have been proposed with most strategies falling into one of two groups: (1) those estimating location sensitivity independently of other locations<sup>4,5</sup> and (2) those who use neighboring locations to estimate sensitivity of related locations.<sup>6–8</sup> The latter category

has received growing interest as it appears to optimize speed and accuracy more effectively.

Most notably, the dynamic test strategy<sup>6</sup> uses a dynamic approach to estimate thresholds at locations and leverages found location values to seed neighboring locations when selected for testing. Alternatively, tendency-oriented strategy (TOP)<sup>8,9</sup> uses neighboring locations to estimate sensitivity thresholds in an asynchronous fashion, leading to extremely fast estimation at the cost of accuracy. More recently, new strategies<sup>10,11</sup> have focused on using neighboring locations in a more data-driven and coherent fashion, which has led to improved performances. Our work follows this line of research as well.

In an effort to further improve perimetry accuracy and speed, we propose in this work an alternative strategy, namely, spatial entropy pursuit (SEP). Our underlying idea is that visual fields should be estimated in a global fashion, where final and intermediate results at different locations can be used to estimate other locations. SEP follows a standard perimetry strategy for individual locations (i.e., zippy estimation by sequential testing [ZEST]) and begins by measuring four locations. Our approach then attempts to locate which locations should be tested next in order to provide the largest information gain with respect to the uncertainty of the entire field. The testing procedure can then be stopped when the overall uncertainty has attained a user-specified level and thus avoids the need to test all locations. In practice, we model the



visual field as a large conditional random field (CRF)<sup>12</sup> represented as a graph of location estimates whose values depend on themselves as well as on their neighbors. Prior location estimates as well as neighborhood relationships are derived from a large data set of glaucomatous visual fields. For comparison, we evaluated the performance of our method against that of existing strategies by using simulations implemented in the open perimetry interface (OPD).<sup>13</sup>

## METHODS

To reduce perimetry examination duration without compromising accuracy, we made use of the spatial relationships between neighboring visual field locations as previously suggested by Chong et al.<sup>10</sup> and Rubinstein et al.<sup>11</sup> Accordingly, an iterative scheme was derived that follows ZEST testing<sup>5</sup> at individual locations and internally determines the next location and stimulus intensity to use by leveraging a visual field model. We begin by explaining our method and the visual field model and how we evaluated our algorithm by using computer simulations.

### ZEST Strategy

Briefly, in ZEST, each visual field location is associated with a prior probability mass function (PMF), which represents the probability of a given sensitivity threshold. ZEST then (1) selects a visual field location randomly from a pool containing all unfinished locations and (2) presents one light stimulus with the stimulus intensity at the mean of the PMF at the selected location; (3) the subject response is incorporated into the PMF in a Bayesian way by multiplying the likelihood function (i.e., probability-of-seeing curve) for the given response with the PMF. Steps (1) to (3) are then repeated until the stopping criterion is reached (i.e., the uncertainty of the PMF, measured by the Shannon entropy,<sup>14</sup> is lower than a predefined value), at which point the location is removed from the pool of unfinished locations. ZEST terminates when all locations have been tested.

To generate the initial PMFs of the visual field locations, we used visual field data of glaucomatous patients from the Rotterdam Ophthalmic Institute's "Longitudinal Glaucomatous Visual Field data"<sup>1,15</sup> (4863 visual fields from 278 eyes of 139 glaucoma patients). To generate smooth PMFs from these data, the PMFs were smoothed by using a Gaussian kernel ( $\sigma = 1.5$ , window = 10 dB).

### Visual Field Model

To infer sensitivity threshold probabilities (i.e., PMFs) for visual field locations from neighboring locations, we used a CRF. In our CRF, a node describes a visual field location and nodes are connected by edges that represent the probabilistic relationships between locations. In this work, we modeled the 24-2 test pattern and assumed that each location (except the ones at the border of the test pattern) has a four-neighborhood connection, as depicted in Figure 1a. Each node corresponds to a PMF representing a visual field location sensitivity threshold probability, and edges are PMFs that encode the conditional probability distributions for pairs of threshold sensitivities (Figs. 1b-d). The edge PMFs are generated from the data and smoothed with a Gaussian kernel ( $\sigma = 2.5$ , window = 10 dB  $\times$  10 dB).

To compute estimates for the unfinished locations we used the Loopy Belief Propagation method,<sup>16</sup> which propagates information found at individual locations to neighboring nodes iteratively and leverages the spatial relationships of locations

encoded via edge connections. In this way, each node influences nodes further away according to probabilistic dependencies. To avoid updating the nodes corresponding to already finished locations, we fixed these nodes' PMFs during the information propagation.

Note that we used a four-neighborhood connection to model location interactions in this work. While simple, other spatial configurations such as retinal nerve fiber layer-inspired connections as described by Jansson et al.,<sup>17,18</sup> could be used instead.

### The SEP Algorithm

The proposed SEP algorithm is an algorithm that uses the ZEST method at given locations but differs in the way the locations to be tested are selected. In particular, we make use of a *pool* of four locations that are being tested at any given point in time and which are initially selected as those with highest uncertainty according to their initial PMF. The algorithm then automatically and dynamically removes terminated locations from this pool and adds new locations in a way that reduces the overall uncertainty of the visual field as much as possible.

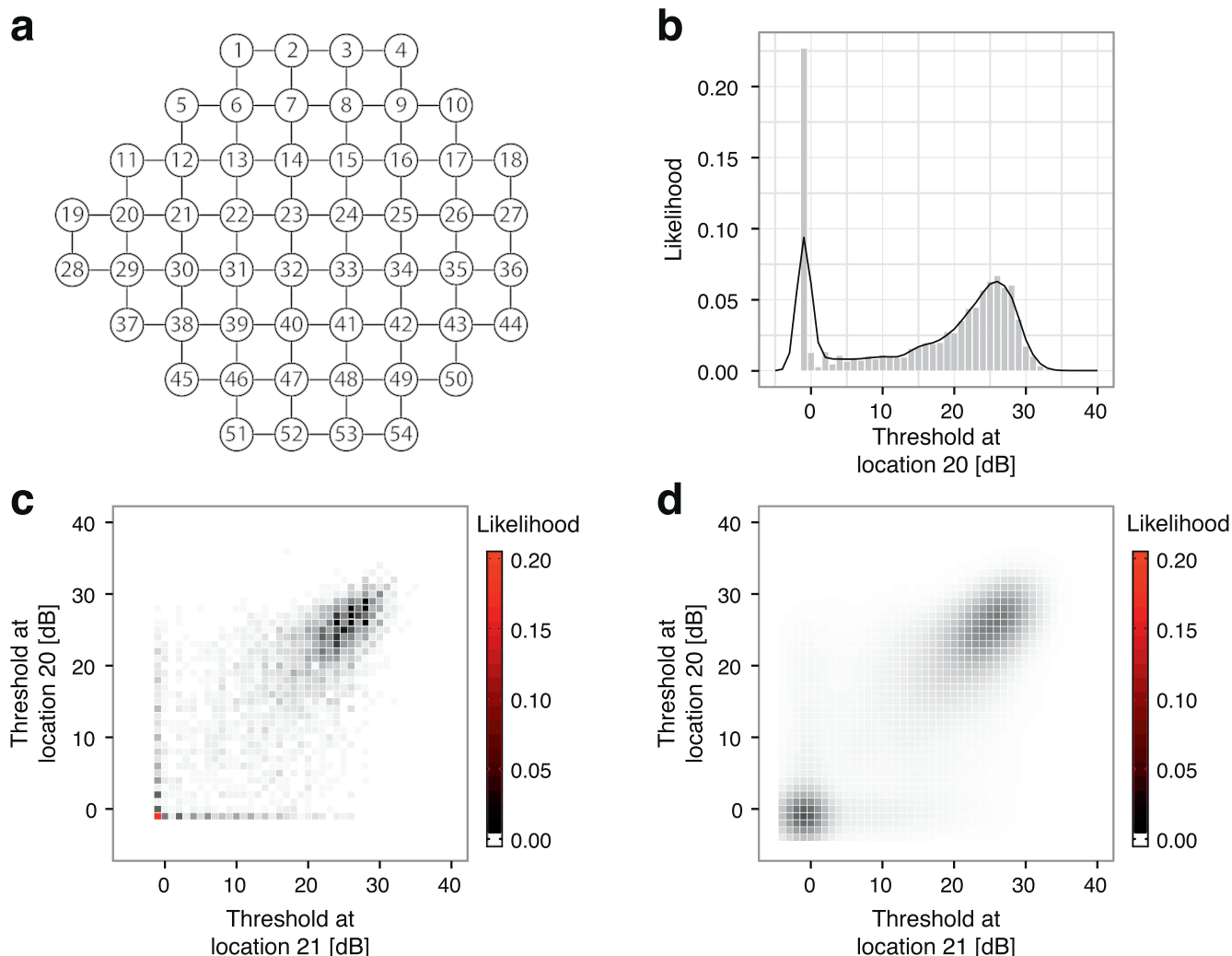
In practice, at each stimuli presentation, one of four locations in the pool is selected randomly and tested by using ZEST as described above. This is done until one of the four locations finishes, at which point it is removed from the pool. A new location is then added in the following way: (1) the current PMFs of all locations are placed into the visual field model; (2) the visual field model is then used to approximate the most likely PMF for each unfinished location as based on all responses over all locations; and (3) PMFs are then used to select the most informative location among the untested locations. To do this we define a function that is computed for each untested location and that combines two measures, namely, the Shannon entropy of the location estimate and its neighborhood heterogeneity,

$$C_i = M_H(i) + \alpha M_G(i), \quad (1)$$

where  $M_H(i)$  and  $M_G(i)$  are nonnegative and stand for the entropy and neighborhood heterogeneity of the visual field estimates at location  $i$ , respectively. The parameter  $\alpha \in \mathbb{R}^+$  is a weight that influences the relative importance of these two factors. The next location is then selected as the one maximizing  $C_i$  in Equation 1, implying that either one or both of the defined measures should be high. The entropy measure,  $M_H(i)$ , quantifies the uncertainty of the modeled PMF of location  $i$ , while the neighborhood heterogeneity,  $M_G(i)$ , represents an approximation of the spatial threshold gradient of the current field estimate (see Appendix A for computational details). In particular,  $M_G(i)$  quantifies neighborhood threshold consistency and is higher at locations whose neighbor estimates differ from one another.

Note that Equation 1 implicitly presumes that poor neighborhood consistency is a good indicator to test. Once the location with the highest  $C_i$  is moved to the pool of locations to be tested, we substitute the PMF of the newly added location with that of the modeled PMF. This is achieved by summing the model PMF with a constant  $\delta$ , to reduce its confidence, and then multiplying it with the prior PMF. This effectively avoids being overconfident in the model probabilities.

With four locations in the pool again, the procedure restarts and continues until  $C_i$  is lower than a predefined value. In this case, no additional location is moved to the pool and the algorithm terminates as soon as the remaining three locations are finished. Importantly, this implies that SEP does not



**FIGURE 1.** Conditional random field model. (a) The graph used for our visual field model (24-2 test pattern). (b) Prior probability mass function of location 20. Bars represent the raw PMF; the line represents the smoothed PMF. (c) Raw PMF for the edge connecting locations 20 and 21. (d) Smoothed PMF for the same edge.

measure all visual field locations and infers sensitivity thresholds for the untested locations from the visual field model after termination of the algorithm.

### Evaluation of SEP

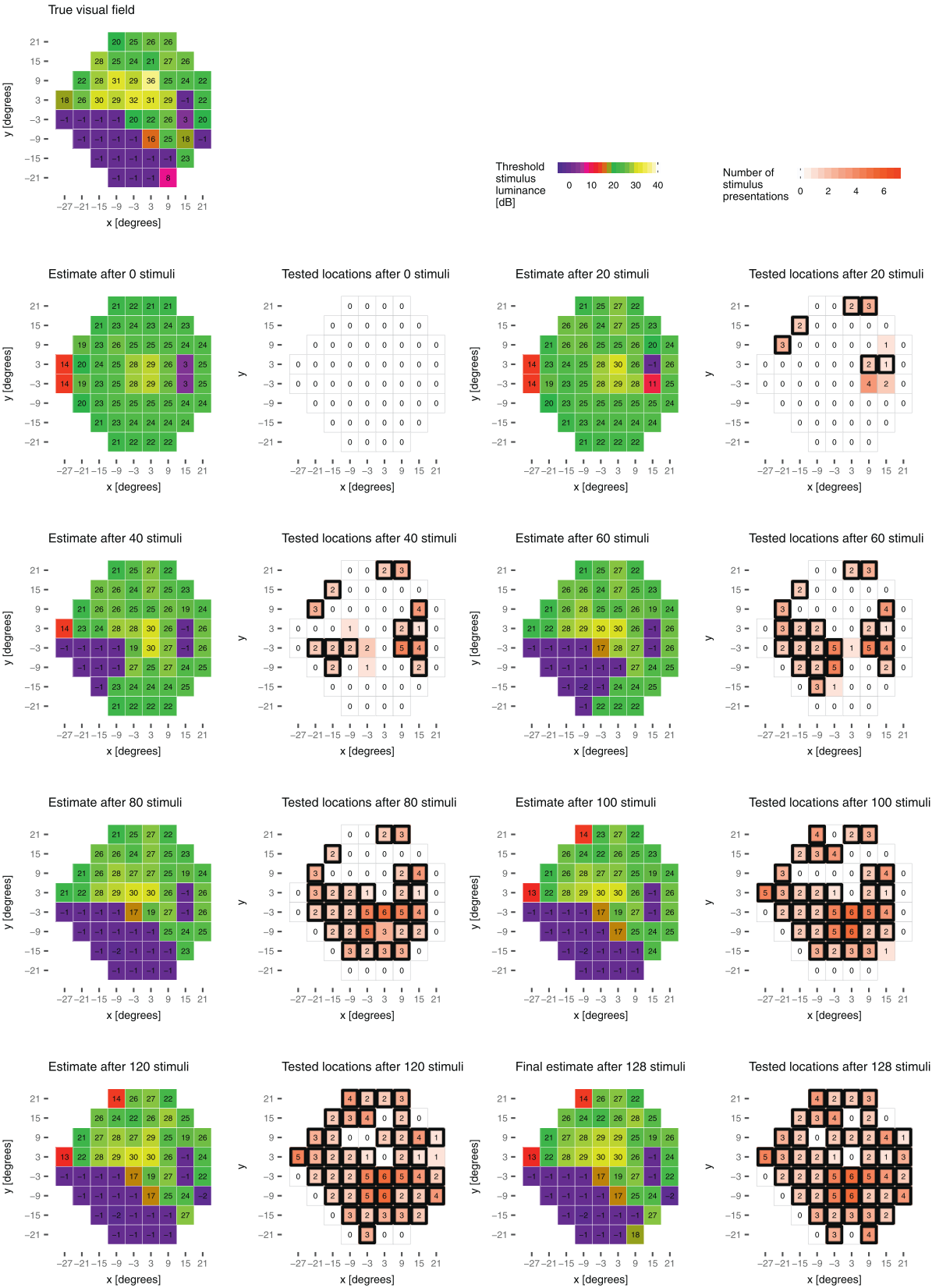
Our algorithm was implemented and evaluated by means of computer simulations using the R package OPI.<sup>13</sup> Responses to stimuli presentations were modeled by sampling from a Frequency-of-Seeing (FOS) curve (i.e., psychometric function) with a predefined false-positive and false-negative response rate of 3% and 1%, respectively. The slope of the FOS curve for a given threshold was modeled with a cumulative normal distribution with standard deviation (SD) according to a published variability formula.<sup>19</sup> The maximum standard deviation allowed for the slope was set to 6 dB. To find suitable parameters for SEP and ZEST (i.e., parameters that minimize the number of stimulus presentations while yielding accuracy levels comparable to the dynamic test strategy), a parameter optimization procedure was performed (see Appendix B) on a subset of the data.

To evaluate SEP's performance in terms of accuracy and testing duration (i.e., number of stimulus presentations), we compared it to ZEST, TOP,<sup>20</sup> and the Haag-Streit dynamic test

strategy, which we denote as ZEST, TOP-like, and dynamic-like, respectively (exact implementations of dynamic test strategy and TOP are not public and may slightly differ from our implementation).

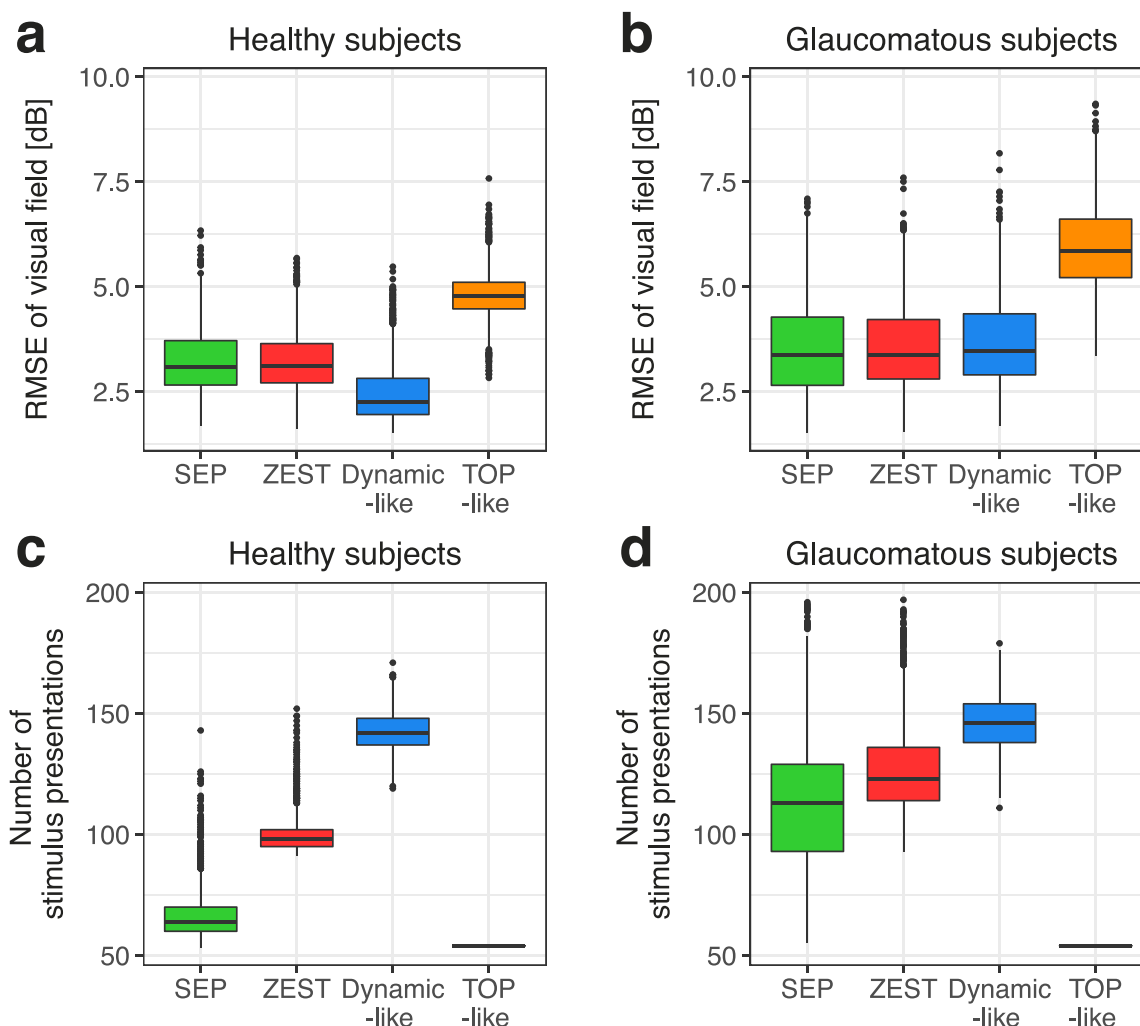
To compare methods, simulations were performed by using a test set with visual fields of 10 randomly selected glaucomatous eyes from the data (on average 17 visual fields per eye [SD = 2.2]), whereby all visual fields of the selected 10 eyes were removed from the data used to generate prior PMFs and edge potentials. Simulations with glaucomatous eyes were performed five times with different test sets, in order to control for selection bias (see Appendix C). In addition, to assess performance for healthy patients, simulations were performed on 245 healthy visual fields (Mean MD = 0.021 dB, SD = 1.7) from the Rotterdam Ophthalmic Institute (control data from the data). Each visual field was measured five times to gather test-retest variability.

To quantify the accuracy of an algorithm, we made use of the RMSE between the true and estimated sensitivity threshold at all locations in a visual field. We also examined how SEP performs on scotoma regions by checking the estimation errors at locations with a high scotoma measure as defined by Rubinstein et al.<sup>11</sup> Here we computed  $max_d$ , by calculating the greatest difference in threshold sensitivity between a location



**FIGURE 2.** Example visual field measurement with SEP. The *top figure* represents the glaucomatous visual field that was measured (MD = -11.8 dB). For each stage, that is, stimulus presentation, the *figure on the left* represents the current visual field estimate and the *figure on the right* shows the number and locations of presented stimuli. *Black boxes* indicate finished locations.





**FIGURE 3.** Performance of SEP compared to existing methods. Number of stimulus presentations and root-mean-square error of visual fields measured with SEP, ZEST, dynamic-like, and TOP-like strategies. Each visual field was tested five times. (a–c) Simulations performed with 245 healthy visual fields (MD ranging from  $-6.4$  to  $3.2$  dB). (b–d) Simulations performed with 172 glaucomatous visual fields (MD ranging from  $-31.1$  to  $5.1$  dB).

and any of its eight adjacent locations (ignoring the blind spot locations 26 and 35). Thus, high  $max_d$  values indicate locations at scotoma borders, whereas low values indicate locations in uniform regions.

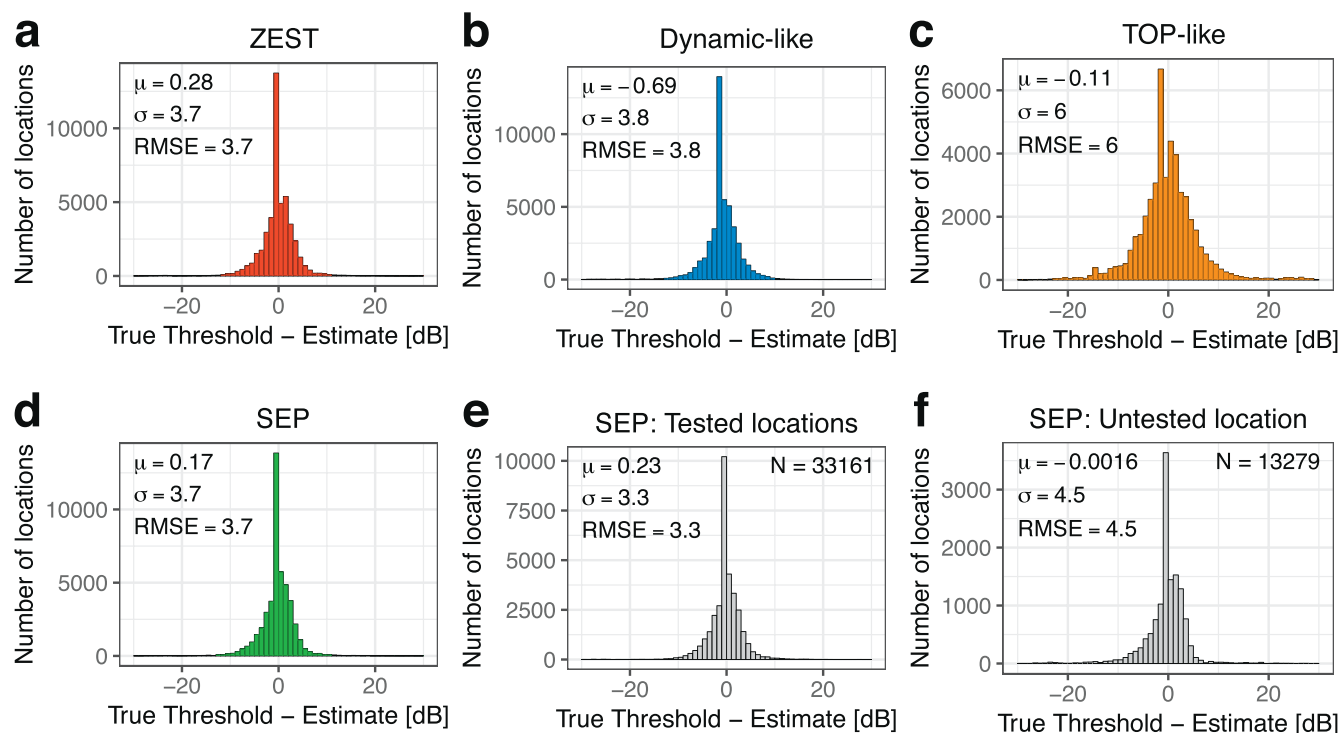
## RESULTS

To provide a qualitative understanding of SEP, Figure 2 depicts the order in which different locations were selected for testing and the associated estimated visual field as function of presented stimuli for a given example. In this example, a glaucomatous visual field with an MD of  $-11.8$  dB was tested with SEP, and the estimated visual field at intervals of stimulus presentations is shown. SEP concluded after 131 presentations, and 11 visual field locations were inferred in this example.

For 245 healthy visual fields, the simulation result is displayed in Figures 3a and 3c. The median RMSE of the tested visual fields was  $3.1$  dB in both SEP and ZEST (nonsignificant differences, Mann-Whitney  $U$  test,  $P > 0.05$ ). Median RMSE in the dynamic-like strategy was  $2.3$  dB (significant difference with SEP, Mann-Whitney  $U$  test,  $P < 0.001$ ). The median number of stimulus presentations was 64 in SEP, 98 in ZEST, and 142 in the dynamic-like strategy (significant differences,

Mann-Whitney  $U$  test,  $P < 0.001$ ). The TOP-like strategy had significantly lower number of stimulus presentations than all other algorithms, but also significantly higher median RMSE of  $4.8$  dB (Mann-Whitney  $U$  test,  $P < 0.001$ ).

Similarly, Figures 3b and 3d report the performances of each method on glaucomatous visual fields (one test set is shown). In this plot the test set contains a total of 172 visual fields from 10 eyes, acquired over an average time span of 8.9 years ( $SD = 1.1$ ). The average MD of the visual fields is  $-10$  dB ( $SD = 7.8$ ). The median RMSE was  $3.4$  dB in both SEP and ZEST (nonsignificant differences, Mann-Whitney  $U$  test,  $P > 0.05$ ) and  $3.5$  dB in the dynamic-like strategy (significant difference with SEP, Mann-Whitney  $U$  test,  $P < 0.01$ ). The median number of stimulus presentations was 113 in SEP, 123 in ZEST, and 146 in the dynamic-like strategy (significant difference, Mann-Whitney  $U$  test,  $P < 0.001$ ). The TOP-like strategy has significantly lower number of stimulus presentations than the other algorithms, but also significantly higher median RMSE of  $5.8$  dB (Mann-Whitney  $U$  test,  $P < 0.001$ ). In SEP, there is no clear dependency of RMSE on MD whereas number of stimulus presentations correlates well with MD especially for MD values between 0 and  $-14$ . This is expected as SEP was optimized as to provide the same accuracy level for any subject by adapting its speed. The relationship



**FIGURE 4.** Pooled errors of all visual field locations. Deviations from true threshold of all visual field locations evaluated during the simulation are shown (a) ZEST, (b) the dynamic-like, (c) the TOP-like strategy, and (d) SEP. For SEP, additionally, (e) the tested and (f) inferred (untested) locations are shown separately.  $N$  is the number of visual field locations. Results are on simulations with glaucomatous visual fields.

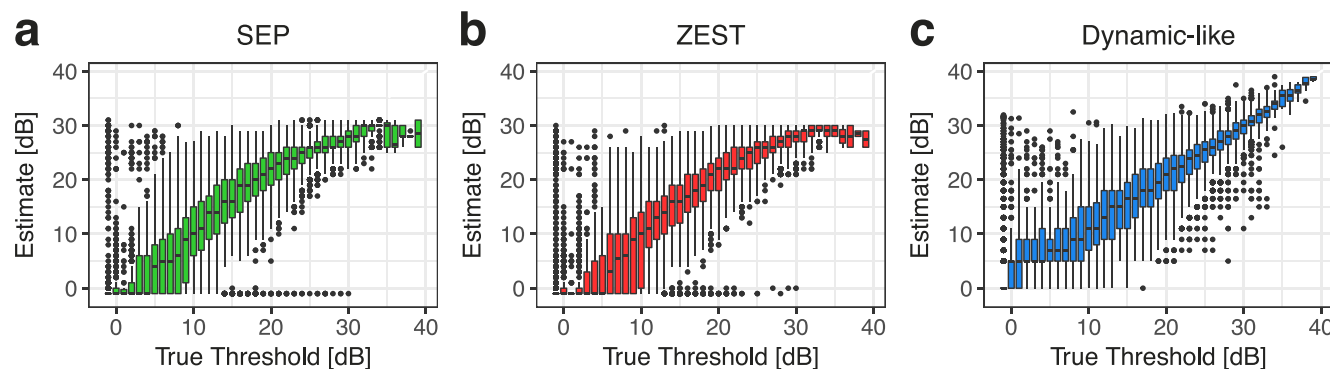
between RMSE and MD however reveals that the median error is highest in visual fields with intermediate MD of  $-14$ , corresponding mostly to heterogeneous visual fields (see Appendix D).

Figure 4 reports the distributions of the pooled errors over all locations and visual fields. The error bias, that is, mean error, was  $-0.11$  dB (SD = 6) in TOP,  $0.17$  dB (SD = 3.7) in SEP,  $0.28$  dB (SD = 3.7) in ZEST, and  $-0.69$  dB (SD = 3.8) in the dynamic-like strategy. The distributions differed significantly between algorithms (two-sample  $t$ -test,  $P < 0.05$ ). Absolute mean error of tested locations of SEP was  $0.23$  dB (SD = 3.3), compared to that of untested locations, which was  $-0.0016$  dB (SD = 4.5). The distributions between tested and untested locations of SEP differed significantly (two-sample  $t$ -test,  $P < 0.001$ ). The average number of tested locations per visual field in SEP was 39 (SD = 5.1) out of 54.

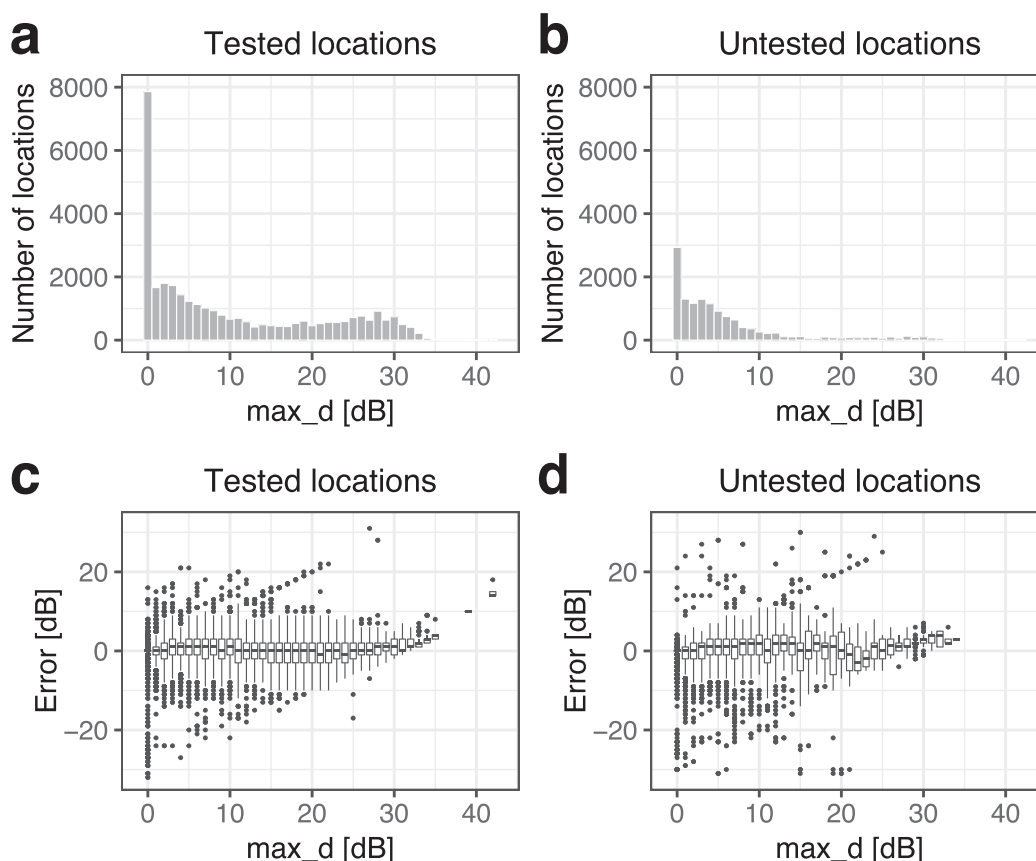
In Figure 5, the algorithm's threshold estimates are grouped as a function of the true threshold, while Figure 6 reports the distribution of  $max_d$  values among (a) tested locations and inferred (untested) locations. Additionally the errors of tested (c) and untested (d) locations are displayed as a function of  $max_d$ .

Figure 7 shows the test-retest variability of SEP (median SD = 1.1 dB), ZEST, and the dynamic-like strategy. SEP retest variability is significantly lower than that of ZEST (median SD = 1.8 dB) and the dynamic-like strategy (median SD = 2.2 dB; Mann-Whitney  $U$  test,  $P < 0.001$ ).

To illustrate the convergence rate of SEP and the impact of the CRF model, we compared ZEST to SEP with no stopping criteria, that is, with all locations tested. After each stimulus presentation, the mean RMSE over all visual fields was calculated and is reported in Figure 8. For comparison, two additional procedures were added: in SEP<sub>entropy</sub>, the next location to test is picked only using the entropy measure ( $\alpha =$



**FIGURE 5.** Accuracy of SEP with respect to threshold sensitivities, compared to existing methods. Median threshold estimate for each underlying true threshold sensitivity. Boxes represent 25th and 75th percentiles and dots depict samples outside of the 1.5 interquartile range. Result based on simulations with glaucomatous visual fields.



**FIGURE 6.** SEP performance at scotoma borders. Number of locations and median measurement error grouped by  $\max_d$  value for tested (a–c) and inferred (untested) locations (b–d). Boxes represent 25th and 75th percentiles and dots depict samples outside of the 1.5 interquartile range. Results are on simulations with glaucomatous visual fields.

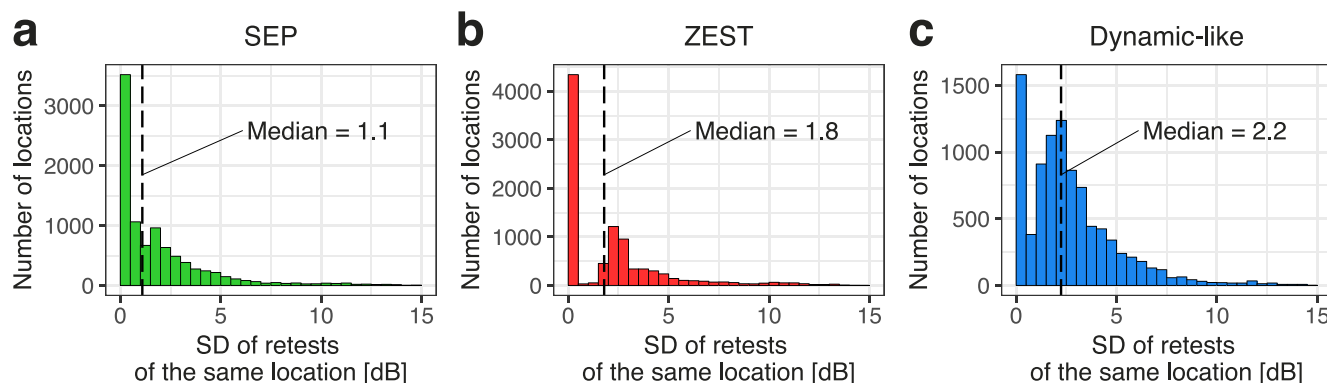
0) and in  $\text{SEP}_{\text{random}}$ , locations are picked randomly. Note that  $\text{SEP}_{\text{random}}$  tests location orders as in ZEST but still infers untested locations by using the CRF. The simulations were performed on a test set of glaucomatous visual fields and each visual field was measured five times.

## DISCUSSION

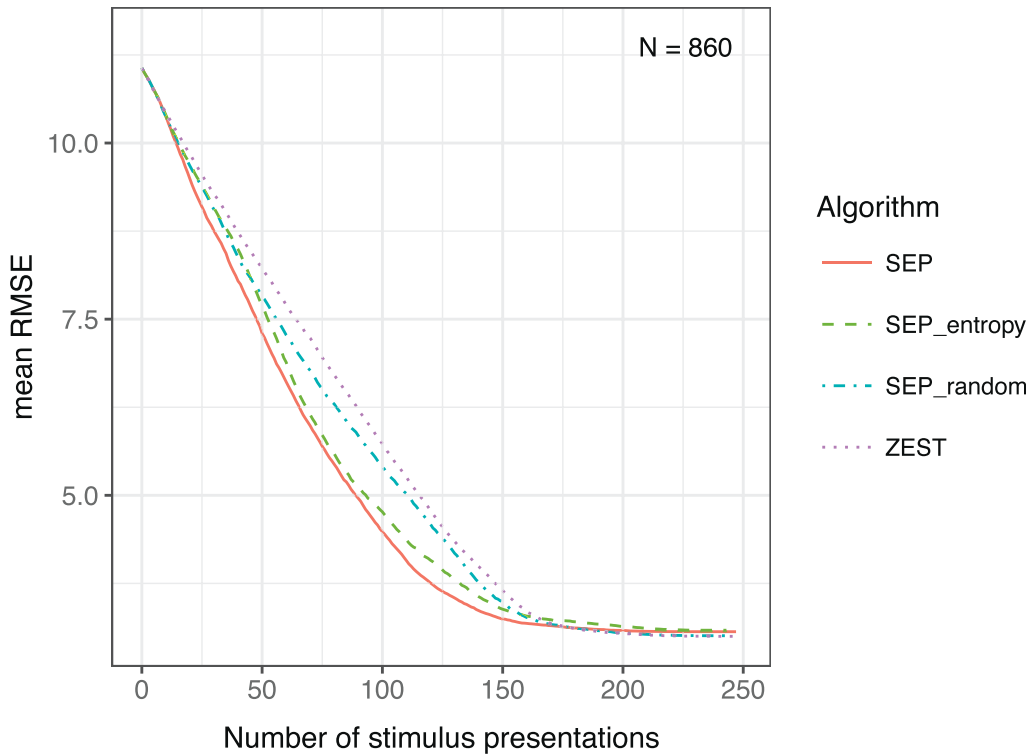
In this work, we presented a novel data-driven approach, namely SEP, for faster perimetry testing without strongly compromising accuracy. SEP leverages existing correlations

between sensitivity thresholds within a neighborhood of locations and assesses which locations need to be tested to reduce the remaining uncertainty as much as possible. We evaluated SEP on a mixed population and compared its performance to state-of-the-art methods.

In the example measurement shown in Figure 2, it can be seen that all locations along the scotoma border are being measured, whereas in more homogeneous regions, within as well as outside of the scotoma, the measurement is coarser, leaving gaps of untested locations. These untested locations are evenly distributed over the homogeneous regions that are ideal for inference with the CRF.



**FIGURE 7.** Test-retest variability of SEP compared to existing methods. Histograms of standard deviations of estimates for the same location are presented for SEP, ZEST, and dynamic-like strategies.



**FIGURE 8.** Error as a function of increasing number of stimulus presentations for different strategies. Mean RMSE of visual fields as a function of stimulus presentations for SEP; SEPentropy, which selects new locations only based on entropy; SEPrandom, which selects new locations randomly; and ZEST. The early stopping was disabled and algorithms terminate only after all locations are measured. Simulations were performed with glaucomatous visual fields. Each visual field was measured five times leading to  $N = 860$  realizations.

Simulations based on data from healthy patients show that SEP can reduce the number of stimuli presentations by 35% when compared to ZEST and by 55% when compared to the dynamic-like strategy (Fig. 3a). While SEP and ZEST roughly have the same accuracy level, the dynamic-like strategy is more accurate for healthy visual fields. SEP and ZEST, however, are still more accurate than any of the tested algorithms when measuring glaucomatous visual fields. With a median number of stimulus presentations of 64, SEP is almost as fast as the TOP-like algorithm for healthy visual fields, while having much lower RMSE. On a wide cohort of early to severe glaucomatous patients, we were able to demonstrate that SEP can reduce the number of stimuli presentations by 8% when compared to ZEST and by 23% when compared to the dynamic-like strategy, while roughly keeping the same accuracy level (Fig. 3b).

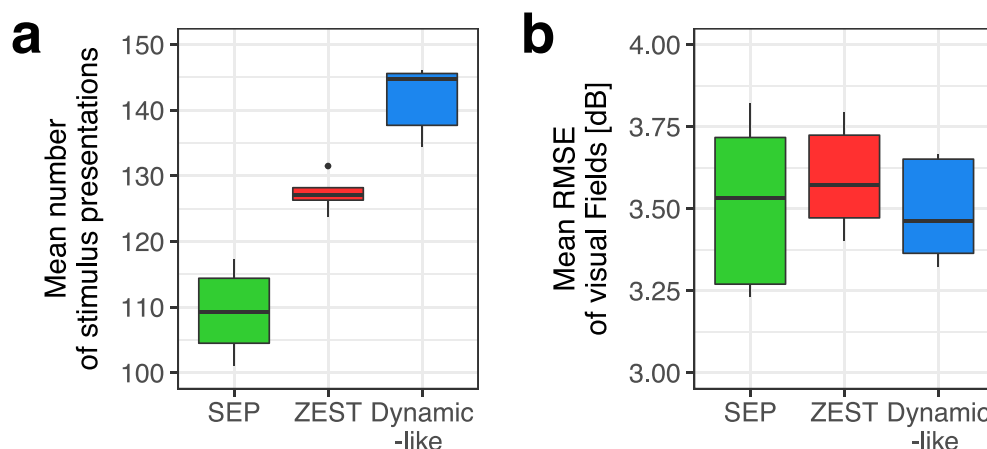
An important difference between SEP and other testing strategies is in the examination speed. In particular, the variance of number of stimuli presentations of SEP is more

than twice as high as that of the dynamic-like strategy for healthy as well as glaucomatous visual fields. This is because SEP does not follow a fixed pattern, but adaptively terminates. In this sense, the speed-up is a result of the reduced number of evaluated locations, which depends directly on the visual field in question. As such, SEP is subject-specific and modulates where it tests accordingly. For healthy visual fields, SEP requires no more than 60 stimuli presentations (see Appendix D), which is almost as fast as TOP, while yielding significantly higher accuracy. For more impaired visual fields, SEP requires approximately 200 stimuli presentations, which is even slightly higher than the dynamic-like strategy. The reason that visual fields with intermediate MD of approximately  $-14$  have the highest median error as well as the highest median number of stimulus presentations (see Appendix D) is attributed to the fact that these visual fields are the least homogeneous and inference, as performed in SEP, is prone to introduce errors.

**TABLE.** Optimal Parameters Used in SEP, Computed in the Parameter Optimization Procedure

Parameter	Value	Description
$update_{fn}$	0.8	False-positive rate in likelihood function for Bayesian update in ZEST
$update_{fn}$	0.8	False-negative rate in likelihood function for Bayesian update in ZEST
$localStopVal$	3.6 (SD = 0.0042)	Limit for entropy at a location to continue testing: test stops if entropy of PMF $\leq localStopVal$
$\sigma_{smooth}^{priors}$	1.5	Standard deviation of Gaussian kernel for smoothing the prior PMFs
$\sigma_{smooth}^{edges}$	2.5	Standard deviation of Gaussian kernel for smoothing the edge PMFs
$nIter$	1	Number of iterations performed by loopy belief propagation algorithm
$\delta$	0.4	Parameter to adjust the confidence we put in the node beliefs from the CRF
$\alpha$	0.05 (SD = 0.0065)	Weight parameter for gradient measure in $C_i$
$globalStoVal$	4.1 (SD = 0.076)	Limit defined for global stopping criterion: SEP ends if $\max C_i \leq globalStoVal$





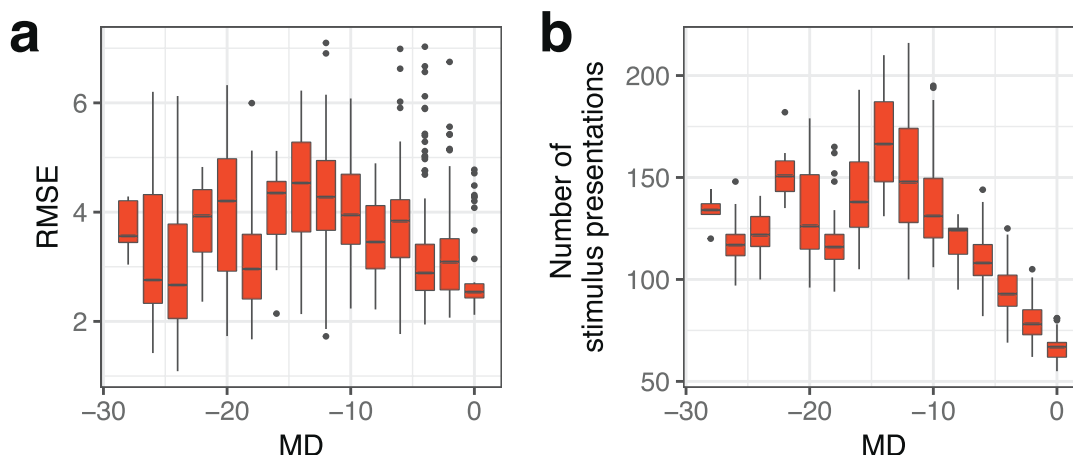
**FIGURE 9.** Cross-validation of parameter sets. Mean performance of SEP for all five parameter sets from the optimization process is presented in terms of number of stimulus presentations (a) and RMSE (b). ZEST and dynamic-like strategy are shown for comparison. Simulations are performed with glaucomatous visual field.

When looking at estimate errors of single locations (Fig. 4), the error distributions of SEP, ZEST, and the dynamic-like strategy are very similar in terms of variance and RMSE. The deviations of the means of the error distributions from zero indicate whether a strategy is systematically biased toward lower or higher values. The errors in the dynamic strategy show the strongest bias, tending to overestimate the thresholds, making them appear healthier than they actually are. SEP and ZEST have smaller bias and tend to slightly underestimate the threshold sensitivity, making locations appear less healthy than they in fact are.

In SEP, some locations are not tested but inferred from a model. While this in theory could lead to large inaccuracies at untested locations, our results showed that this is in fact not the case (Fig. 4f). The mean error of 0 dB indicates that the inferred thresholds are not biased toward higher or lower values. The distribution is in fact slightly skewed and in some rare cases the thresholds are being overestimated by as much as 17 dB. As can be seen by the lower error SD, SEP has overall fewer cases of over- and underestimation than ZEST and the dynamic-like strategy. The error SD of the untested locations of SEP is slightly higher than that of measured locations in SEP, ZEST, and the dynamic-like strategy, which is to be expected. It is, however, only 18% higher than in the dynamic-like algorithm where the locations are measured, and still 25% lower than in the TOP-like strategy. Indeed, the relatively low

error rate in untested locations can most likely be accounted for by our CRF model that propagates information in a coherent and appropriate fashion.<sup>21</sup> This can be interpreted as smart smoothing technique where SEP does not allow information propagation through the already tested locations, which avoids smoothing thresholds at the locations that have been thoroughly tested.

Even though SEP and ZEST, on average, have a lower bias than the dynamic-like strategy, it can be seen from the relationship between true threshold and estimated threshold (Fig. 5), that for some thresholds SEP and ZEST are biased as well. For thresholds of 4 dB and lower, SEP and ZEST tend to underestimate threshold sensitivities, whereas the dynamic-like strategy tends to overestimate threshold sensitivities. For true threshold values of 29 dB and higher, SEP and ZEST show a tendency to underestimate threshold sensitivities, not allowing values higher than 31 dB. While underestimation of the visual field is most likely preferred over overestimation, it highlights a limitation of our approach. The reason for this bias lies in the prior probability distribution used for the individual visual field locations. The threshold estimates are clearly biased toward the most likely values of -1 dB and values of approximately 27 dB shown in Figure 1a. This is also supported by existing evidence in the literature<sup>22</sup> as well as by the fact that if a uniform PMF were used, this behavior would not be observed. This effect is stronger when using a smaller  $\sigma$  for the Gaussian



**FIGURE 10.** Performance dependency of SEP on MD. RMSE (a) and number of stimulus (b) presentations as a function of the visual fields MD are presented.

smoothing kernel (i.e., the peak at  $-1$  is more pronounced) and can be prevented by using a higher  $\sigma$ , but this has shown to increase the acquisition time significantly. As such, research in a more appropriate approach to model the prior probabilities could further improve performances for visual fields with high variability.

It can be seen in Figure 6b that among untested locations, barely any locations have a  $max_d$  value higher than 12 dB. This indicates that most locations at scotoma borders (high  $max_d$ ) are in fact being tested. Looking at the errors, it can be seen that for tested locations the error generally is not higher at scotoma borders where  $max_d$  is high. Among untested locations, the error is generally higher than in tested locations (as already observed in Fig. 4f), but without an error increase as a function of  $max_d$ . This indicates that for rare cases where a location with high  $max_d$  remains unmeasured, these do not significantly increase the error in our algorithm.

To our surprise, the test-retest variability (Fig. 7) of SEP was the lowest among all tested algorithms including ZEST. This was unexpected since it is unlikely that two measurements of the same visual field with SEP measure the exact same visual field locations. The low test-retest variability can, however, be accounted for by the fact that in SEP the tested locations are measured with higher precision than the locations in ZEST, as can be seen in Figures 4a and 4e.

When looking at the visual field RMSE of intermediate steps of SEP, ZEST, and related algorithms, as shown in Figure 8, it can be seen that SEP decreases RMSE much faster than ZEST. Thus, if stopped early, SEP has lower error after the same number of stimulus presentations than ZEST. In addition, it can be seen that SEP is slower in decreasing error when locations are picked by using only the entropy ( $SEP_{entropy}$ ). Lastly, it can be seen that ZEST eliminates errors faster when using a CRF to get intermediate estimates for untested locations ( $SEP_{random}$ ). This is to be expected since the estimates for untested locations in ZEST are normative values that can be uninformative in some cases.

A potential limitation of our algorithm is the high dependency of SEP on the model parameters. The automatic parameter optimization procedure we present helps mitigate the challenge in selecting many codependent parameters, but our solution is clearly nonoptimal. Better fine-tuning is likely to lead to improved performances. Additionally, the parameter optimization step could be further refined by targeting different pathologic subpopulations. How the selected parameters would affect results on different subpopulations remains an open question though.

Another limitation of SEP is that it requires significant amounts of data to model the neighborhood graph, which in its current form only takes the location-wise closeness into account. Alternative edge connectivity between test locations could be modeled as well, and in particular more anatomically and physiologically plausible models would be appropriate as proposed by Rubinstein et al.<sup>11</sup> Further analysis would be helpful to better understand the influence of the neighborhood model on the performance.

In summary, we introduced a new adaptive perimetry testing strategy that exploits spatial information within the visual fields in order to estimate threshold sensitivities with high precision and fewer number of stimulus presentations. We proposed to model visual fields with a CRF, which helps leverage neighboring information. During an examination, SEP decides on the location and the stimulus to query at the next step and stops when overall confidence in the visual field estimate has been reached. It thus infers threshold sensitivities of untested locations, which notably decreases the number of stimuli presentations needed. With an appropriate selection of parameters, SEP has been shown to provide the same accuracy

with less stimuli presentations than ZEST and the dynamic-like strategy on glaucomatous visual fields. In the future, we will investigate the performance of this approach on human subjects in order to show its clinical relevance.

## Acknowledgments

The authors thank Hans Bebie and Stefan Zysset for their comments and help with our dynamic-like strategy implementation.

Disclosure: **D. Wild**, None; **S.S. Kucur**, Haag-Streit Foundation (F); **R. Sznitman**, Haag-Streit Foundation (F)

## References

1. Bryan SR, Vermeer KA, Eilers PHC, Lemij HG, Lesaffre EM. Robust and censored modeling and prediction of progression in glaucomatous visual fields. *Invest Ophthalmol Vis Sci*. 2013;54:6694-6700.
2. Zulauf M, Flammer J, LeBlanc RP. Normal visual fields measured with octopus program G1. *Graefes Arch Clin Exp Ophthalmol*. 1994;32:509-515.
3. Li SG, Spaeth GL, Scimeca HA, Schatz NJ, Saving PJ. Clinical experiences with the use of an automated perimeter (otopus) in the diagnosis and management of patients with glaucoma and neurologic diseases. *Ophthalmology*. 1979;86:1302-1312.
4. Flanagan JG, Wild JM, Trope GE. Evaluation of fastpac, a new strategy for threshold estimation with the humphrey field analyzer, in a glaucomatous population. *Ophthalmology*. 1993;100:949-954.
5. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res*. 1994;34:885-912.
6. Weber J, Klimaschka T. Test time and efficiency of the dynamic strategy in glaucoma perimetry. *Ger J Ophthalmol*. 1995;4:25-31.
7. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997;75:368-375.
8. Gonzalez de la Rosa M, Martinez A, Mesa Sanchez C, Cordoves L, Losada MJ. Accuracy of tendency-oriented perimetry with the octopus 1-2-3 perimeter. In: *Perimetry Update, 1996/1997: Proceedings of the XIIth International Perimetric Society Meeting*. Vol 1997. Würzburg, Germany: Kugler Publications; 1996:119-123.
9. González de la Rosa M, Bron A, Morales J, Sponsel WE. Top perimetry (a theoretical evaluation) (abstract). *Vision Res Sup Jermov*. 1996;36:88.
10. Chong LX, McKendrick AM, Ganeshrao SB, Turpin A. Customized, automated stimulus location choice for assessment of visual field defects introduction of goanna. *Invest Ophthalmol Vis Sci*. 2014;55:3265-3274.
11. Rubinstein NJ, McKendrick AM, Turpin A. Incorporating spatial models in visual field test procedures. *Transl Vis Sci Technol*. 2016;5:7.
12. He XR, Zemel RS, Carreira-Perpiñán MA. Multiscale conditional random fields for image labeling. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington DC: IEEE Computer Society; 2004:695-702.
13. Turpin A, Artes PH, McKendrick AM. The open perimetry interface: an enabling tool for clinical visual psychophysics. *J Vis*. 2012;12(11):22.
14. Shannon C. A mathematical theory of communication. *Bell System Tech J*. 1948;27:79-423.
15. Erler NS, Bryan SR, Eilers PHC, et al. Optimizing structure-function relationship by maximizing correspondence be-

- tween glaucomatous visual fields and mathematical retinal nerve fiber models optimization of structural RNFL models on visual fields. *Invest Ophthalmol Vis Sci*. 2014;55:2350-2357.
16. Murphy KP, Weiss Y, Jordan MI. Loopy belief propagation for approximate inference: an empirical study. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers Inc. 1999:467-475.
  17. Jansonius N, Nevalainen J, Selig B, et al. A mathematical description of nerve fiber bundle trajectories and their variability in the human retina. *Vision Res*. 2009;49:2157-2163.
  18. Jansonius N, Schiefer J, Nevalainen J, Paetzold J, Schiefer U. A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position. *Exp Eye Res*. 2012;102:70-78.
  19. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41:417-421.
  20. Anderson AJ. Spatial resolution of the tendency-oriented perimetry algorithm. *Invest Ophthalmol Vis Sci*. 2003;44:1962-1968.
  21. Olsson J, Rootzén H. An image model for quantal response analysis in perimetry. *Scand J Stat*. 1994;21:375-387.
  22. Mckendrick AM, Turpin A. Combining perimetric suprathreshold and threshold procedures to reduce measurement variability in areas of visual field loss. *Optom Vis Sci*. 2005;82:43-51.
  23. Gonzalez RC, Woods RE. *Digital Image Processing*. 2nd ed. Boston, MA: Addison-Wesley Longman Publishing Co., Inc. 2001.

## APPENDIX A

### Neighborhood Heterogeneity

The neighborhood heterogeneity,  $M_G(i)$  is computed from the current visual field estimate  $\mathbf{E}$  constructed from the CRF visual field estimate. Let  $\mathbf{E}$  be an  $8 \times 9$  matrix,

$$\mathbf{E}x = \begin{bmatrix} 0 & 0 & 0 & x'_1 & \cdots & 0 & 0 \\ 0 & 0 & x'_5 & x'_6 & \cdots & 0 & 0 \\ 0 & x'_{11} & x'_{12} & x'_{13} & \cdots & x'_{18} & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & x'_{51} & \cdots & 0 & 0 \end{bmatrix},$$

where  $x'$  represents the median of the PMF at location  $i$  and matrix entries that do not correspond to a visual field location are padded with zeros as in the study by Gonzalez and Woods.<sup>23</sup>  $M_G(i)$  is computed by using two  $3 \times 3$  kernels, which are convolved with the visual field estimate  $\mathbf{E}$  to get approximations of the horizontal and vertical derivatives, respectively.  $M_G(i)$  can then be computed as

$$M_G(i) = \sqrt{G_x^2 + G_y^2} \quad (2)$$

$$\text{with: } G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{E} \quad \text{and} \\ G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \mathbf{E}.$$

## APPENDIX B

### Parameter Selection and Optimization

Given the different parameters our strategy uses (Table), we now describe an automatic strategy to establish valid parameters from a data set of visual fields. As mentioned in Section Evaluation of SEP, our data set is divided into two: a training and a test set. The test set contains visual fields of 10 randomly selected eyes from the data set. The rest of the visual fields are assigned to the training set. The training set is used to generate prior PMFs as well as edge potentials for the CRF.

In addition, a small part of the training set is also used to establish the parameters of our approach. This is done in two steps. First, 5% of the visual fields in the training set are randomly selected and used to optimize the parameters such as  $update_{fp}$ ,  $update_{fn}$ ,  $localStopVal$ ,  $\sigma_{smooth}^{priors}$  (Table). Simulations are performed with different values for each parameter. The parameter configuration was chosen so as to minimize the mean acquisition error, that is, RMSE while not exceeding the mean number of stimuli presentations of dynamic-like test strategy. Second, another 5% of the visual fields in the training set are randomly selected and used to optimize the parameters such as  $\sigma_{smooth}^{edges}$ ,  $nIter$ ,  $\alpha$ ,  $\delta$ , and  $globalStopVal$  (Table) while fixing the above parameters at the found values from the first step. The parameters that minimized the mean number of stimuli presentations while not exceeding the mean acquisition error of dynamic-like test strategy were chosen.

To avoid selection bias in this procedure, we repeated this process five times by using different permutations of training and test sets. This yields five different parameter sets for our algorithm. We report simulation results using each of these parameter sets in Appendix C, while the Table shows means and standard deviations of the parameter sets used in this work.

## APPENDIX C

### Results: Cross-Validation

To demonstrate robustness of our automatic parameter optimization process, a 5-fold cross-validation process was performed. We optimized the parameters, using the training sets, and evaluated SEP five times, using different disjoint training-test set pairs (see Appendix B). We report in Figure 9 the performances (mean RMSD versus mean number of steps) over the simulations, using each of the five parameter sets and their respective test set. For comparison, the performances of ZEST and the dynamic-like strategy on the same five test sets are shown as well. The median of the mean number of stimulus presentations for SEP was 109, for ZEST 127, and for the dynamic-like strategy 145. The median of the mean RMSE of SEP was 3.53 dB, compared to 3.57 dB for ZEST and 3.47 dB for the dynamic-like strategy.

## APPENDIX D

### Results: Dependency of Performance on MD

In Figure 10, we show the dependency of SEP performance on the MD of the visual fields.